

Exercise 3

1. For the first question you have to use Martview, so go to this web page
2. Choose the database
3. Choose the dataset (Human)
4. In the Filter section, select Region, Chromosome 10, and type 12000000 and 14000000 in Gene Start and Gene End
5. To limit your search to genes that are expressed in the liver, in the Expression section, select "eGenetics/SANBI EST anatomical system data" and type `liver` or get to liver from the Browse button.
6. Click Results
7. For the other questions you have to go to MapViewer, latest Build, Homo sapiens
8. Click on Chromosome 10 and then type 12000000 (or 12M) and 14000000 (or 14M) in the Region Shown window.
9. In the "Maps and Options" section, add STS (microsatellites) and Variations (SNPs). You can remove everything else.

Exercise 4

Q1: Is this genomic sequence or cDNA (RNA?)

cDNA. You can find this out by simply blasting one of the sequences

Q2: What gene these sequences belong to?

CHLORIDE CHANNEL 1, again, the answer comes from BLAST.

Q3: Is this gene a "disease-gene"?

Yes, Congenital myotonia. You can find this by looking at the "links" hyperlink of the gene, under OMIM or PubMed, etc.

Q4: These 4 sequences are not identical. Where do they differ?

A:

```
AATCCACACAGCTGACTGGCCCAGGCTCTCAGGATCACCCG[t]GTGTTTCA
CCCATGTATTGTTGTCAAACAAAGTACTGATGGCTTCGCGGGGCATCAACT
CTCCAGCCATGAATTG
```

B:

```
AATCCACACAGCTGACTGGCCCAGGCTCTCAGGATCACCCGCGTGTTTCA
CCCATGTATTGTTGTCAAACAAAGTACTGATGGCTTCGCGGGGCATCAACT
CTCCAGCCATG[a]ATTG
```

C:

```
AATCCACACAGCTGACTGGCCCAGGCTCTCAGGATCACCCGCGTGTTTCA
C[t]CATGTATTGTTGTCAAACAAAGTACTGATGGCTTCGCGGGGCATCAAC
TCTCCAGCCATGAATTG
```

D:

AATCCACACAGCTGACTGGCCCAGGCTCTCAGGATCACCCGCGTGTTTCA
CCC[t]TGTATTGTTGTCAAACAAAGTACTGATGGCTTCGCGGGGCATCAAC
TCTCCAGCCATGAATTG

Q5: Can you describe the type of DNA change for each of these sequences?

A: polymorphism. Ala437Thr

B: mutation. Phe413Cys

C: not described, probably mutation. Trp433End

D: not described, probably polymorphism. Thr432Thr

BONUS QUESTION

Do a CTRL+F search on the genomic part of Ensembl

Exercise 5

1. Go to the EntrezGene web site and type CFTR. Select CFTR in human. Scroll down until you find the reference sequence for its mRNA (NM_000492.3) [The same result can be achieved by starting from the ENSEMBL web site]
2. Copy the mRNA sequence
3. Go to the ORFfinder web site (via NCBI tools), paste the sequence, and detect all possible ORFs in this mRNA
4. Note that the first ORF (sense) is the longest one, so it is likely the good one. However, this same ORF has a very long region on its 3' that does not seem to be coding, so you want to check whether everything is in order.
5. Click on this ORF and on "accept"
6. Click on View Fasta nucleotide or Fasta protein (it's the same)
7. Go to the CFTR page in ENSEMBL, select CFTR-001 ("001" usually corresponds to the "canonical transcript), and get to its cDNA.
8. Compare the protein or the nucleotide that you obtained with ORF finder and the protein or mRNA present in ENSEMBL (for example with bl2seq --see exercise No.2--). The sequences are identical.